SHIP CLASSIFICATION IN HIGH AND VERY HIGH RESOLUTION SATELLITE SAR IMAGERY

Virginia Fernandez Arguedas[1] Domenico Velotto[2], Björn Tings[2], Harm Greidanus[1] and Carlos Bentes[2]

[1] *harm.greidanus@jrc.ec.europa.eu*
European Commission - Joint Research Centre (JRC),
Via E. Fermi 2749, 21027 Ispra (Italy)

[2] *domenico.velotto@dlr.de*
Maritime Safety and Security Lab, Remote Sensing Technology Institute,
German Aerospace Center (DLR), Henrich-Focke-Str. 4, 28199, Bremen (Germany)

Abstract

To serve the security of the maritime domain, ship self-reporting systems provide information on the cooperative vessels. However, non-reporting ships should be also monitored. Satellite images can be used to detect and classify non-reporting ships. Synthetic Aperture Radar (SAR) offers monitoring capabilities regardless of clouds or daylight, and hence it is used for satellite global monitoring. Different satellite SAR systems are deployed, from European ones such as Sentinel-1, to national ones such as TerraSAR-X, presenting very diverse characteristics from their coverage to their image resolution. In this paper, two ship classification methods are presented, a method developed for use on high (20 m) resolution SAR images (Sentinel-1 dataset), and a method developed for use on very high (3 m) resolution ones (TerraSAR-X dataset). In a cross-application experiment, both methods are evaluated on both datasets. The exercise quantifies the methods' performance across resolutions, highlighting their pros and cons in this challenging application.

Keywords: Maritime Surveillance, Ship Classification, Synthetic Aperture Radar.

1 INTRODUCTION

To ensure adequate security of the maritime domain, including the maritime borders, it is necessary to be aware of the shipping activities in the relevant areas of sea. Some areas are covered by shore-based monitoring and observation systems, but for those that are not, satellite-based systems are an efficient alternative. The two main tools for satellite-based maritime surveillance are (a) automatic ship reporting systems and (b) imaging systems. Automatic ship reporting systems such as AIS (Automatic Identification System), LRIT (Long Range Identification and Tracking) or VMS (Vessel Monitoring System) let ships report their identity and position on a regular basis, and are mandated for certain classes of ships by specific (national or international) regulations [e.g., 1]. LRIT and VMS are restricted to government use, while satellite AIS is more widely available on a commercial basis. Data from these systems, and in particular from AIS that covers most ships in the world of 300 gross tonnes and up, enable the tracking of most of the medium and large ships globally. However, not all the AIS-carrying ships are successfully seen by satellite due to noise and interference problems, and ships engaged in irregular activities might turn off their AIS to avoid attention. Moreover, most of the smaller ships do not use automatic ship reporting systems.

In order to also find non-reporting ships out of coastal sensor range, satellite imaging is used, in optical or radar frequencies. The first step in analysing such images for maritime surveillance is ship detection (finding the ships), the second step is ship classification (establishing the type of ship). Although optical images are more readily interpreted for classification, they are hampered by clouds and are not available at night. Radar performs regardless of clouds or daylight and is therefore often preferred for maritime surveillance. Synthetic Aperture Radar (SAR) is the type of radar used from satellite, synthesising the radar image during the few seconds that the satellite illuminates the target.

The signature of a ship in a SAR image is, however, not easily interpreted, making classification difficult. It depends on the detailed shape of the metal structures on the ship, appearing differently depending on viewing geometry. The ship's signature can merge with radar backscatter from the ship's immediate surroundings, and its motions on the waves during the radar illumination time can introduce a blurring. Therefore, classification based on the ship's signature in SAR images is a challenge.

Satellite SAR systems can provide images with a range of spatial resolutions, from over 100 meter to below 1 meter. While the high resolutions are obviously preferred for their better classification power, the image sizes are inversely proportional to resolutions, low resolution images covering up to 450 km as opposed to only 5 km at the very high resolution end. Wide-area maritime surveillance can therefore not be done at very high resolution.

Europe has several satellite SAR systems in operation. Among those, the European Union's Copernicus program offers the Sentinel-1 SAR [2], which provides daily routine coverage of many maritime areas including the European seas. The most frequently produced image type, suitable for maritime surveillance, is the Interferometric Wide (IW) mode Ground Range Detected High resolution (GRDH) product that has 250 km swath width at 20 m resolution.

Among the national systems, the German TerraSAR-X [3] can produce images on demand with swaths between 270 km and 5 km and resolutions, respectively, between 40 m and 0.25 m. A good compromise between coverage and resolution is achieved by the Stripmap imaging mode with a 30 km swath and 3 m resolution (Multi-look Ground range Detected, MGD, product).

This paper discusses approaches to classify ship signatures in images from these two SARs. Although the minimum detectable ship size is an ill defined concept because it depends very strongly on ship type, radar parameters, viewing geometry and ambient conditions, as a rough indicator half of the resolution can be taken. In relation to the use of AIS, the 300 tonnes limit very roughly corresponds to 45 m ship length. For Sentinel-1 (IW-GRDH product), the size ranges of relevance are therefore (a) 10 m to 45 m, for the smaller ships that are still detectable but do mostly not report on AIS; and (b) 45-400 m, for the medium and large ships that are subject to the use of AIS. At 20 m resolution, however, even the large ships do not show much detail. It is therefore too much asked to do a classification into all relevant ship types, which would include passenger ship, tanker, container, bulk carrier, fishing ship, patrol ship, tug, etc. Instead, the ambition to perform classification into any possible ship type is reduced to a classification problem into a restricted set of classes. In many maritime areas of interest, the most frequently occurring types are tanker and cargo. It is still useful to be able to distinguish between those two. The ship reporting data contains the ship type (cargo or tanker), so a disambiguation between those two types in the SAR targets allows deciding on a probable association between a known reporting ship and a target found in the SAR image. Therefore, the classification problem for the Sentinel-1 images is here reduced to a disambiguation between two ship types, cargo and tanker.

For TerraSAR-X, having more resolving power, the classification problem is generalised to maritime object classification adding three further classes of maritime targets: offshore platform, offshore wind turbine and harbour structure.

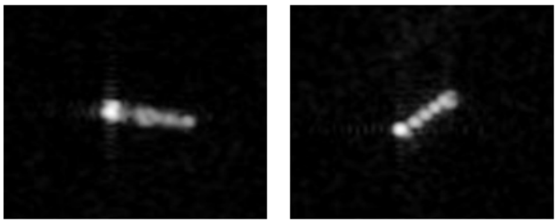
Having available the two classification schemes, the 2-class disambiguation developed for the lower resolution Sentinel-1 images and the 5-class classification developed for the higher resolution TerraSAR-X images, it is also attempted to do a cross-application. The paper will present results of applying each of the two algorithms to each of the two data sets.

2 DATA

2.1 Sentinel-1 and related reference data

Four Sentinel-1 IW-GRDH images over the Western Indian Ocean were obtained during a period where also AIS data from up to 17 satellites operated by four providers were collected. The providers were exactEarth, SpaceQuest, ORBCOMM/LuxSpace and the Norwegian Coastal Administration. The SAR images were subjected to ship detection with JRC's SUMO detector, which uses a Constant False Alarm Rate (CFAR) algorithm, resulting in a total of 146 targets. Only co-pol (HH and VV) channels were used, without making a distinction between the two. After visual verification of the targets, correlation with AIS ship positions, and selecting only those AIS ships that were unambiguously tankers and cargo ships, a total of 100 targets were retained, 71 cargo ships and 29 tankers. Image chips of 140x140 pixels were extracted around the targets for the classification, and spatially upsampled by a factor of two, to 5 m pixel spacing.

Cargo



Tanker

Figure 1. Examples of cargo and tanker classes from the Sentinel-1 dataset.

2.2 TerraSAR-X and related reference data

A total of 75 TerraSAR-X Stripmap MGD images, mostly over the North and Baltic Seas, were acquired in areas that were covered by terrestrial AIS and where known clusters of platforms and wind parks exist. As for the Sentinel-1 data, only co-pol channels were used without distinction. While the full range of incidence angles accessible to the TerraSAR-X Stripmap mode is 20° to 45°, a single such image covers a much smaller range of incidence angles that a single Sentinel-1 IW image. The relatively high number of acquisitions was necessary to capture the variations over incidence angle and marine conditions. The ship detector used for the SAR image dataset was DLR's SAINT detector, also of the CFAR type. The detected targets are automatically collocated with AIS and platform / wind park position databases. If no match has been found, the SAR detected target is discarded from the classification dataset. At the end of this process a total of 683 targets were extracted, representing the 5 classes of interest, see Fig. 2 [4].

Ship Windpark Oil platform Harbor

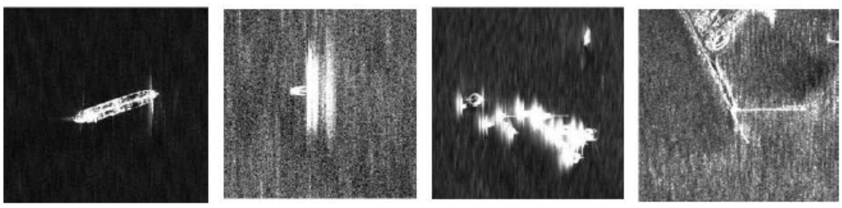


Figure 2. Examples of target classes form the very high resolution TerraSAR-X data set.

Additionally, a subset containing 185 cargo and tanker ship type (distribution 135 cargo and 50 tankers) was extracted to pursue the disambiguation problem. Image chips of 128 x 128 pixels were extracted around the targets and resampled to a common 2.5 m pixel spacing.

3 CLASSIFICATION METHOD

Two different methods for classification were applied to the data; one that was designed for the lower resolution Sentinel-1 data, and one that was designed for the very high resolution TerraSAR-X data.

3.1 Classification method designed for Sentinel-1 images

The image chips around the SAR signatures are first subjected to pre-processing to separate the (brighter) ship signature from the surrounding (darker) background. This consists of edge detection and morphological operations to delineate a contiguous signature outline while discarding isolated bright pixels. The image chip is then rotated so that the signature outline long axis points horizontally, and all pixels outside the smallest rectangular box that contains the signature outline are removed. The result is a reduced, rectangular, horizontal image chip that just fits the ship signature. The reduced chip is split lengthwise into three parts that represent bow, middle and stern.

Two texture measures are computed, Local Binary Patterns (LBP) and Histogram of Gradients (HOG). LBP analyses the immediate neighbours of each pixel, considering local spatial patterns and grey scale contrast [5, 6]. LBP is computed for the entire reduced chip and also for the bow, middle and stern parts separately. HOG calculates the distribution of intensity gradients or edge directions [7]. HOG is computed only for the entire reduced chip.

The classification has two phases, training and testing. In the training phase, a training dataset is analysed to build up a dictionary composed of a representative set of feature samples from each of the two vessel classes, tanker and cargo ship, based on Bag of Visual Words [8]. In the testing phase, the extracted features for each sample are compared with the built-in dictionary to determine the vessel class by finding the nearest neighbour. The general structure of the system is based on the work presented in [9,10]. For each of the two ship types, 15 samples were used for the training. All 100 samples were used for testing.

Each of the classifiers (LBP overall, LBP bow, LBP middle, LBP stern, HOG overall) gives a certain percentage result for correct disambiguation between the two classes cargo ship and tanker. The disambiguation result can be further improved by combining several classifiers. Several combinations using major voting as fusion method were tried out to find the best combination.

3.2 Classification method designed for TerraSAR-X images

Also here, the image chips extracted from the satellite images are first subjected to some pre-processing steps. The details can be found in [4] and consist in radiometric calibration, removing the ocean clutter signature, isolating the target of interest in the chip and normalising the intensity of the target's signal response. The classification dataset is artificially enlarged performing a set of label-preserving transformations. The resulting augmented dataset is composed of 500 samples per class providing a more balanced classification dataset. For the training step 90% of the data are used and the remaining 10% are used for testing.

The classification model proposed here is based on Deep Neural Network (DNN). In [4] different DNN architectures have been analysed and the initial results encourages the use of Convolutional Neural Network (CNN) for the maritime object classification problem. The initial results obtained using high resolution TerraSAR-X images show that with an ensemble of CNN models an average f1-score of 93% for the considered 5 classes of maritime objects is achieved [4]. The advantage of DNN classifiers is the possibility to learn complex non-linear problems without the need of extracting handcrafted class features. However, the training process might be computationally expensive and an optimal network setting needs to be found. Here we briefly introduce to the final model architecture and topology developed in [4].

Fig. 3 shows the graph representation of the CNN model used. The input fed to the network is the SAR image chip obtained after pre-processing. The connections between neurons inside the network are achieved by a convolution operator and optimised to work with images. The proposed CNN topology is composed of two convolutional layers, alternated by pooling layers in order to reduce the dimensionality, and a fully connected dense (D) layer followed by a softmax (S) layer with dimensions provided in Fig. 3.

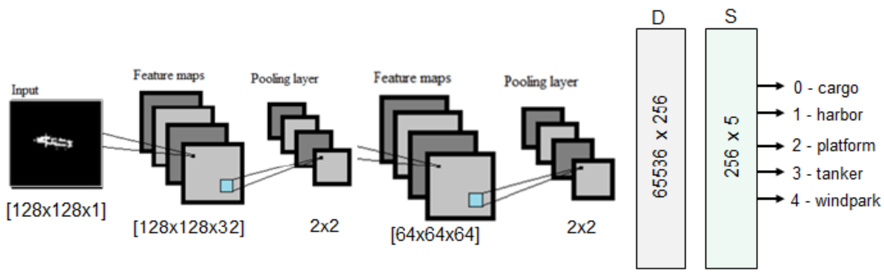


Figure 3. CNN architecture. The input is the chip obtained after pre-processing. The output is the class labels.

4 RESULTS

The classification performances are reported in terms of precision, recall and f1-score.

4.1 Classification results for the Sentinel-1 method

4.1.1 Sentinel-1 method on Sentinel-1 data

A pool of features was proposed to perform ship classification (disambiguation) for Sentinel-1 data in Section 3.1. Averaged over the testing dataset, the five individual classifiers (LBP overall, bow, middle, and stern, and HOG overall) scored 83 % in precision, 69 % in recall and 75 % in f1 for cargo, while for tanker the classifiers decreased their performance achieving 51 % in precision, 70 % in recall and 59 % in f1. The outperforming features were LBP bow, BP stern and HOG. When evaluated separately for cargo ships and for tankers, the scores differ by between 1 to 11 percentage points.

Table 1 summarises the results for combined classifiers. Three combinations are used, being equally-weighted linear combinations of: (1) General LBP, LBP Bow, LBP Middle and LBP Stern; (2) General LBP and HOG; and (3) LBP Bow, LBP Middle, LBP Stern and HOG. As shown in the table, the results obtained by the combined classifiers clearly exceed the individual classifiers, maximising the individual classifiers' complementary information. The results show that Combination 3 gives the best results achieving a 79.0 % in precision, a 77.9 % in recall and a 78.4 % in f1.

4.1.2 Sentinel-1 method on TerraSAR-X data

The same methodology presented in Section 3.1 is applied to TerraSAR-dataset. The underlying idea is to study the performance of the Sentinel-1 method on a more detailed dataset. The method, built for the analysis of low-resolution images, exploits mainly the texture characteristics of the targets since such features tend to be present in low resolution images. However, it neglects local features or key points detectors since such characteristics are not likely to be present in low resolution images.

Table 1. Sentinel-1 method on Sentinel-1 data.

	Combination 1 %			Combination 2 %			Combination 3 %		
	Precision	Recall	f1	Precision	Recall	f1	Precision	Recall	f1
Cargo	85.0	76.1	80.3	80.9	82.1	81.5	85.5	88.1	86.8
Tanker	57.9	71.0	63.8	60.0	58.1	59.0	72.4	67.7	70.0
Avg/Total	71.5	73.5	72.0	70.4	70.1	70.3	79.0	77.9	78.4

Averaged over the testing dataset, the five individual classifiers scored 76 % in precision, 73 % in recall and 77 % in f1 for cargo, while for tanker the classifiers decreased their performance achieving 35 % in precision, 39 % in recall and 36 % in f1. Amongst all individual classifiers, the feature that outperforms the others is HOG achieving 74 % and 43 % in f1 for cargo and tanker respectively. Comparing HOG performance in Sentinel-1 and TerraSAR-X datasets, its performance clearly improves for TerraSAR-X dataset, obtaining the highest tanker representativity.

Table 2 summarises the results for the Sentinel-1 method on the TerraSAR-X dataset. As in the results obtained in Section 4.1.1, Combination 3 presents a good performance when compared with the individual classifiers. However, for TerraSAR-X data, Combination 1 and 2 outperform Combination 3.

Table 2. Sentinel-1 method on TerraSAR-X data.

	Combination 1 %			Combination 2 %			Combination 3 %		
	Precision	Recall	f1	Precision	Recall	f1	Precision	Recall	f1
Cargo	79.2	84.4	81.7	78.1	92.6	84.8	75.8	90.4	82.4
Tanker	48.8	40.0	44.0	60.0	30.0	40.0	45.8	22.0	29.7
Avg/Total	64.0	62.2	62.8	69.1	61.3	62.4	60.8	56.2	56.1

The obtained results reveal that the selected features performed better on the Sentinel-1 dataset, maintaining performance on the TerraSAR-X dataset for the cargo class but presenting a performance drop on the tanker class. The features selected for the Sentinel-1 method neglect details and representative points due to absence of those in the Sentinel-1 dataset and focus on general appearance, texture and edges. Considering the obtained results, other features focusing more on key points and local information might perform better on higher resolution TerraSAR-X data. This could be further addressed in the future.

4.2 Classification result for TerraSAR-X method

4.2.1 TerraSAR-X method on TerraSAR-X data

Table 3 summarises the results for the 5-class problem obtained by an ensemble model built training CNN architectures in Fig. 3 with input SAR image chips at different pixel spacing. It is important to note that the scores have been obtained using only the test dataset (10% of the overall classification dataset) in order to have more reliable performance estimation.

Table 3. CNN model on TerraSAR-X data. Score obtained for each class.

	Precision %	Recall %	f1-score %
Cargo	100	92	96
Harbor	81	90	85
Platform	97	100	98
Tanker	98	97	97
Windpark	89	85	87
avg / total	93	93	93

4.2.2 *TerraSAR-X method on Sentinel-1data*

Table 4 summarises the results for the disambiguation problem (cargo ship - tanker differentiation) using the CNN model on Sentinel-1.

Table 4. CNN model on Sentinel-1 data. Score obtained for the disambiguation classes.

	Precision %	Recall %	f1-score %
Cargo	86	62	72
Tanker	45	76	56
avg / total	74	66	68

These results have been obtained using the Sentinel-1 dataset described in the section 2.1 directly as test set to the model previously trained on TerraSAR-X images. In this sense can be thought as an experiment of "Transfer Learning". The option to train a model using only Sentinel-1 data will be conducted when a larger classification dataset will be collected and is therefore left for future work.

5 SUMMARY AND CONCLUSION

In this paper, two ship classification methods were presented and evaluated over two datasets with different image resolutions. One method is feature-based, developed on the lower resolution (20 m) Sentinel-1 IW images; the other is image-based, developed on the higher resolution (3 m) TerraSAR-X Stripmap images. The Sentinel-1 method exploits general appearance features, texture and edges, neglecting local features and details. The TerraSAR-X method uses CNN for the classification, avoiding the extraction of handcrafted class features and building the model directly from the images. In a cross-application test for the sub-problem of disambiguation between cargo ship or tanker (two-class classification), it was found that each method performs best on the data for which it was designed. This is attributed to the fact that they exploit their inherent benefits, i.e., the TerraSAR-X method was trained directly on its images, which is possible due to their high quality, while the Sentinel-1 method uses feature-based analysis to compensate for the lower image resolution. However, both methods present promising results in this preliminary experiment.

The cargo-tanker disambiguation with Sentinel-1 data is successful, despite the fact that the Sentinel-1 SAR IW signatures for cargo and tanker ships look quite similar by eye. Nonetheless, the TerraSAR-X data have much more potential for classification than the Sentinel-1 data. A classification into 5 classes that is possible with TerraSAR-X was not attempted with Sentinel-1.

In the future, a larger image dataset will be built and further experiments will be conducted. Moreover, multi-class ship classifiers will be targeted in an attempt to contribute to increased security in the maritime domain.

ACKNOWLEDGMENTS

Sentinel-1 data are © Copernicus 2014-2016.

TerraSAR-X data were provided by the German Aerospace Center (DLR) via AO accounts OCE2917 and OCE1045.

AIS data from Norwegian satellites were obtained courtesy of the Norwegian Defence Research establishment (FFI) and the Norwegian Coastal Administration under the PMAR project.

REFERENCES

[1] Greidanus, H., Vespe, M. and Alvarez, M. (2016). *Detection of anomalous behaviour in ship reporting data for improved maritime security*. Proc. 11th Future Security Conference, Berlin, 13-14 Sep 2016, Fraunhofer Group for Defence and Security VVS (these proceedings).

[2] European Space Agency (2013). *Sentinel-1 User Handbook*. GMES-S1OP-EOPG-TN-13-0001, 1 Sep 2013.

[3] Eineder, M., Fritz, T., Mittermayer, J., Roth, A., Boerner, E., and Breit, H. (2008). *TerraSAR-X Ground Segment, Basic Product Specification Document*. Cluster Applied Remote Sensing (CAF), Oberpfaffenhofen, Germany. Report No. TX-GS-DD-3302.

[4] Bentes, C. Velotto, D. and Lehner, S. (2015). *Target classification in oceanographic SAR images with deep neural networks: Architecture and initial results*. Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, 26-31 Jul 2015, pp. 3703-3706.

[5] Ojala, T., Pietikäinen, M. and Harwood D. (1994). *Performance evaluation of texture measures with classification based on Kullback discrimination of distributions*. Proc. 12th IAPR International Conference on Pattern Recognition, Jerusalem, 9-13 Oct 1994, Vol. 1, pp. 582-585.

[6] Ojala, T., Pietikäinen, M. and Harwood D. (1996). *A Comparative Study of Texture Measures with Classification Based on Featured Distributions*. Pattern Recognition, Vol. 29, pp. 51-59.

[7] Dalal, N. and Triggs, B. (2005). *Histograms of Oriented Gradients for Human Detection*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, 20-25 Jun 2005, Vol. 1, pp. 886-893.

[8] Sivic, J. and Zisserman, A. (2003). *Video Google: A text retrieval approach to object matching in videos*. Proc. 9th IEEE International Conference on Computer Vision, Nice, 13-16 Oct 2003.

[9] Fernandez Arguedas, V. (2015). *Texture-based vessels classifier for electro-optical satellite imagery*. Proc. IEEE International Conference on Image Processing (ICIP), Quebec, 27-30 Sep 2015.

[10] Santamaria, C., Stasolla, M., Fernandez Arguedas, V., Argentieri, P., Alvarez, M. and Greidanus, H. (2015). *Sentinel-1 Maritime Surveillance: Testing and Experiences with Long-term Monitoring*. JRC98532, EUR 27591 EN, ISBN 978-92-79-53960-2 (PDF), ISSN 1831-9424 (online), doi: 10.2788/090400, Luxembourg: Publications Office of the European Union.